

DTIC FILE COPY

①

A REVIEW OF PERSONALITY MEASUREMENT IN AIRCREW SELECTION

D. L. Dolgin and G. D. Gibb

AD-A200 392

DTIC
ELECTE
NOV 16 1988
S H D

88 11 16 047

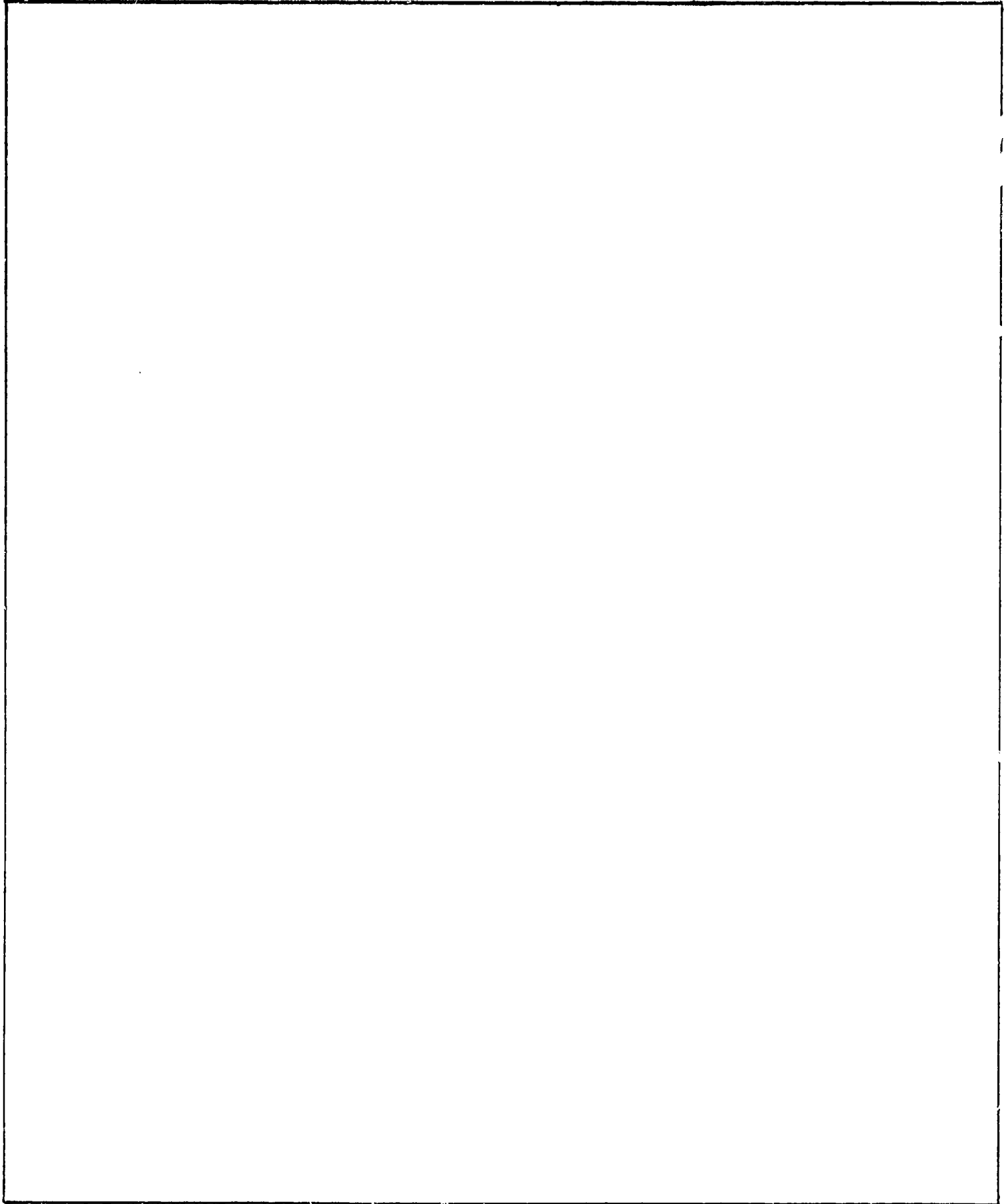
Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, Florida 32508-5700

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY Unclassified			3. DISTRIBUTION / AVAILABILITY OF REPORT		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NAMRL Monograph 36			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Naval Aerospace Medical Research Laboratory		6b. OFFICE SYMBOL (If applicable) 03	7a. NAME OF MONITORING ORGANIZATION Naval Medical Research and Development Command		
6c. ADDRESS (City, State, and ZIP Code) NAS, Pensacola, FL 32508-5700			7b. ADDRESS (City, State, and ZIP Code) NMC, NCR, Bethesda, MD 20814-5044		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63706N	PROJECT NO. M0095001	TASK NO. 001
11. TITLE (Include Security Classification) (U) A Review of Personality in Aviation Selection					
12. PERSONAL AUTHOR(S) D.L. Dolgin and G.D. Gibb					
13a. TYPE OF REPORT Monograph		13b. TIME COVERED FROM 1987 TO 1988		14. DATE OF REPORT (Year, Month, Day) July 1988	
15. PAGE COUNT 40					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Personality, selection, performance measurement, computer-based testing. (5000)g		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) A comprehensive review of personality literature as it relates to aircrew selection was conducted. The purpose of the study was to identify tests that warrant further research as potential prediction instruments. The advent of performance-based personality assessment and implications for future test development were examined. The majority of personality tests reviewed were invalid for pilot selection. Several tests appear to be both effective in pilot selection and psychometrically sound, and we recommend continued research of those. These recommended selection tests include the Defense Mechanism Test, the Personality Research Form, and the Strong Vocational Interest Blank. <i>Keywords: Flight training, Personal assessment,</i>					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL J.A. BRADY, CAPT, MSC, USN, Commanding Officer			22b. TELEPHONE (Include Area Code) (904) 452-3286		22c. OFFICE SYMBOL 00

SECURITY CLASSIFICATION OF THIS PAGE



ABSTRACT

A comprehensive review of personality literature as it relates to aircrew selection was conducted. The purpose was to identify tests that warrant further research as potential prediction instruments. The advent of performance-based personality assessment and implications for future test development were examined. The majority of personality tests reviewed were invalid for pilot selection. Several tests appear to be both effective in pilot selection and psychometrically sound. These personality tests include the Defense Mechanism Test, the Personality Research Form, and the Strong Vocational Inventory Blank.



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Acknowledgments

The authors would like to dedicate this report to the memory of Captain James O. Houghton, MC, USN. Without his idea for developing the monograph, his constant encouragement and support of the work, this study and related efforts could not have been accomplished.

We also wish to acknowledge the help of two individuals, Dr. John de Lorge and Kathleen S. Mayer, who both read several versions of the manuscript and made many helpful suggestions, most of which were incorporated into this work.

INTRODUCTION

Attrition from military flight training is costly in monetary as well as human terms. Each flight student who attrites from the jet training program represents a loss of \$804,783 to the Navy (1). Since World War I, military psychologists have tried to reduce attrition by developing valid tests to select candidates who will complete training programs and continue on as aviators. The aviator selection devices in use today, which primarily assess aptitude, have a validity correlation of approximately 0.15 to 0.25 to a pass/fail criterion for undergraduate pilot training (2). Because aptitude testing alone cannot predict all failures, personality variables and decision-making styles that will improve the selection process become more critical.

Our objective was to explore personality factors used to predict performance in aviation. We use the American Psychiatric Association's definition of personality: "The characteristic way in which a person thinks, feels, and behaves; the ingrained pattern of behavior that each person evolves, both consciously and unconsciously, as the style of life or way of being in adapting to the environment" (3, p. 103). We would like to emphasize that these behavior patterns are relatively stable throughout an individual's life, barring highly unusual circumstances. This is an important underlying assumption in any discussion of personality testing, for we must assume that a personality measure administered at a given point in time is a reliable reflection of the degree of the particular trait that we are attempting to measure.

HISTORICAL INFORMATION

Before World War II, selection for aviator training in the military was based primarily on physical qualifications, with minimal criteria, and the desire to be a pilot or an aircrew member. As the United States entered into the War, the military needed to select large numbers of men in a manner that was cost effective, efficient and, ultimately, safe. Because so many personnel were needed, desire and interest were no longer feasible requisites for aviator selection as many applicants did not possess the skills needed to complete the rigorous academic ground school and preflight aspects. Thus, selection programs evolved to predict those who could complete flight training (4). Consequently, the military has based selection for aviator training on paper-and-pencil performance test batteries since World War II.

Both World Wars I and II catalyzed the development of applied psychology. World War I was the first opportunity for psychologists to test large numbers of applicants, which led to many advances in "intelligence testing" and "mental testing" in the 1920s and 1930s. When World War II started, psychologists had already acquired sufficient test experience and data to apply their techniques to more specific attributes than "intelligence." In this context, aviation provided fertile ground for test development (5), although less rigorous and systematic efforts were attempted earlier (6).

Unannounced Justification	
By _____	
Distribution/ Availability Codes	
Dist	Avail and/or Special

A-1

Test development in aviation selection evolved into four general areas of individual differences assessment: general intellectual measures, aviation-related paper-and-pencil measures, psychomotor performance measures, and personality measures (7-9). These areas have varying degrees of utility in selection and receive different emphasis in Navy and Air Force selection procedures.

The Early Years

The Army Air Forces Aviation Psychology Program conducted a comprehensive investigation of the use of personality measures to predict aviation performance (10). The thrust of the effort was to determine the predictive value of a number of commercially available tests. A secondary consideration was to use questionnaire items from these tests to establish a pool of items of high predictive value in aviation screening. Although performance measures in an actual combat environment were desirable criteria, they were not obtainable. The criterion used for the validation efforts was graduation/ elimination from primary flight training. These studies are summarized in Table 1.

With very few exceptions, personality measures did not predict success in primary flight training. Given the vast number of dependent measures that could be extracted from the personal and preference inventories and their subscales, several measures should have achieved significance by chance factors alone. In addition, item-validation analyses failed to produce many questionnaire items with statistically significant validities. Further, no data were presented to indicate whether any of the measures that reached statistical significance explained any additional variance beyond that accounted for by the existing selection system. Guilford (10) attributed the failure to predict success in flight training to three factors: (1) the tests were not designed to predict flight performance, (2) motivational factors compensated for weakness in personality traits during training, and (3) subject biases yielded inaccurate measures of the personality trait under study.

Clinical evaluations derived from several observations and interviews produced similar results. Clinical ratings based on subjective evaluation were "consistently ineffective in the prediction of success or failure in primary flying training" (13, p. 669). However, a number of methodological problems were inherent in this effort. With respect to the clinical evaluators: (1) No effort was made to control for variability in skill or experience, (2) subjective weightings of the personality dimensions of interest were not uniform, and (3) data were inadequate to assess inter-rater reliability. Clinical evaluations were not used in combination to assess a pilot candidate's chances for success, nor were any other criteria used other than graduation/elimination in primary flight training.

Literature Reviews

Ellis and Conrad (11) summarized the personality literature from 1932 to 1948, which assessed the validity of 26 personality inventories in military practice and included 94 studies on pilots and navigators. Twenty of the studies used aircrew members as the sample population with the following 10 personality inventories: Personal Inventory, MMPI, Bernreuter Personality Inventory, Humm-Wadsworth Inventory, Information Blank, Minnesota Personality Scale, Personal Audit, Inventory of Factors GAMIN, Inventory of STDCR, and the Guilford-Martin Personnel Inventory. Two types of criteria were used:

TABLE 1. Results of Validation Studies.

Personality measure	Sample size (student pilots)	Predictive validity
Information blank	200	None
Humm-Wadsworth temperament scale	202	Hysteroid scale $r=-.19$, $p=.05$ Epileptoid scale $r=-.22$, $p=.05$
Adams-Lepley personal audit scale	271	None
Bernreuter personality inventory	600 graduates & 200 attrites from primary training	None
Inventory of factors STDCR (introversion/extroversion)	1100	None
Guilford-Martin personnel inventory	950	Objectivity scale $r=.10$, $p=.05$ Agreeableness scale $r=.12$, $p=.01$ Cooperative scale $r=.14$, $p=.01$
Inventory of factors GAMIN	780	None
Minnesota multiphasic personality inventory	856	None
Minnesota personality scale, male form	338	None
Shipley personal inventory, format B	1419	None
Restricted word association test	NA	Validation not conducted
Strong vocational interest blank for men	650	None
Maller-Glaser interests values inventory	524	Economic scale $r=.15$, $p=.02$
Kuder preference record	937	None
Teacher preference scale	422	Social sensitivity scale $r=-.16$, $p=.05$
Rorshach-individual administration	156	None
Rorshach-group administration picture exercises test	591	Popular responses score $r=.21$, $p=.01$ Percent animal responses score $r=.14$, $p=.05$ Rejection score $r=-.14$, $p=.05$
Visualization multiple choice	811	None
Thematic apperception test: 38 category scoring	293	None
20 category scoring	191	None
Rapid projection test adapted from Murray rapid projection slides	556	None
Empathetic response test	1028	None
Observational/interview techniques	170 minimum per method	None

psychiatric evaluations and performance. Of seventy studies utilizing psychiatric criteria, 67 reported favorable results. Generally, unfavorable results were obtained when personality inventories were validated against performance criteria. Ellis and Conrad (11) attributed the lack of success to the following reasons:

1. Pre-selection of candidates eliminated abnormal individuals.
2. Performance measures were unreliable and invalid.
3. Individual differences in performance depended more on differences in aptitude and previous training than on any differences in personality.
4. Personality inventories were originally validated against a psychiatric criterion and not against performance measures.

The authors concluded that personality inventories demonstrate little promise in the prediction of performance.

North and Griffin (6) reviewed aviator-selection literature from 1917 to 1977. These authors found that at least 40 different personality inventories and scales were evaluated for pilot selection between 1950 and 1976 "without any appreciable impact on the selection of aviator candidates" (6, p. 18). Only a few studies that examined the use of personality testing to predict voluntary withdrawal from flight training achieved any success. Those investigations that were successful generally added very little predictive power to existing models, were not cross-validated, or failed to cross-validate. Griffin and Mosko (15) attributed the lack of success primarily to test-response bias. All studies that they reviewed involved the selection of naval aviation candidates, a group which they contend are highly susceptible to response faking because of the quality of the candidate pool: 1) all had college degrees; 2) as a group, all were above average in intelligence; and 3) all were highly motivated and sensitive to the effect of performance data on their continuity in a flight program. These characteristics also contribute to a lack of variability among group members (see methodological problems).

Sells (12) reviewed the literature on personality tests used for the selection of flight personnel. Of the 100 tests evaluated, 26 had significant validity coefficients ranging from $r = .10$ to $.45$. Motivational factors, such as attempting to make a good impression, were considered to have an .PA impact on the predictive validities. Overall, four areas demonstrated the highest potential:

1. Aviation Interest Key ($r = .37$ to $.41$ with the pass/fail criterion).
2. The following MMPI scales: a) hypochondriasis, b) psychopathic deviate, c) neuroticism, d) manifest anxiety, e) antisocial, f) depression, and g) hysteria. Significant correlations ranged from $.10$ to $.35$ with pass/fail.
3. Pilot Opinionnaire (evaluates attitudes toward military aviation) correlated $.28$ with pass/fail.

4. Daily grade slips (forms for instructor ratings), which contained instructor comments regarding students' reactions in-flight. Correlations with pass/fail for cadets ($N = 384$) and officers ($N = 66$) were .36 and .58, respectively, for number of comments by instructors; .35 and .56 for number of comment categories; .32 and .56 for the average daily grade, and .39 and .64, respectively, for the composite of all three scores. The information from daily grade reports of the first 10 flights provided an important predictor of training outcome after a brief period of actual flight instruction.

The Personal Inventory, Cornell Index, Cornell Word Form, and the School of Aviation Medicine Sentence Completion Test were validated against post-training operational and combat criteria. The results yielded low correlations ranging from .04 to .23.

The Navy has studied aviator personality and performance (see 6 and 9 for reviews) to determine which candidates are not motivated to complete training. Traditional tests, such as the Minnesota Multiphasic Personality Inventory (MMPI) and the Taylor Manifest Anxiety Scale, do not consistently provide unique predictive validity (13). One reason is that they are designed to detect psychopathology rather than specific performance (14). Similarly, tests developed to assess "normal personalities" (e.g., the California Psychological Inventory) also have little value in predicting success in aviation training (15).

Specific Test Research

Minnesota Multiphasic Personality Inventory (MMPI). The MMPI is the most widely used personality test (over 3,500 references published; 16). It consists of 550 statements to which the subject responds either true, false, or cannot say. The MMPI provides measures on 10 clinical scales: hypochondriasis, depression, hysteria, psychopathic personality, masculinity-femininity, paranoia, psychasthenia, schizophrenia, hypomania, and social introversion. It was developed by Hathaway and McKinley (17) to diagnose psychopathology. Compared to other personality tests used in aviation, the MMPI generally has been the most successful in predicting training success.

Melton (18) found that specific combinations of MMPI scales, rather than individual scale scores, were related to success in flight training. Subjects with low scores on hysteria (Hy), masculinity-femininity (Mf), and mania (Ma) were in the "flight failure" category. Conversely, the "flight completion" group was defined by high Hy, Mf, and Ma scores. A discriminant function for the two clusters resulted in no overlap. Melton correctly classified 83% of a Navy cadet sample population into pass/fail categories based on MMPI scores.

In another study, Fulkerson et al. (19) used the MMPI to determine the appropriateness of the test's norms on a pilot population ($N = 634$); the validity of the individual scales and the validity of the K-correction (a measure of defensiveness of test-taking attitude). They found that the norms for the pilot sample differed significantly from that of the original normative group. The MMPI did not differentiate significantly between pass/fail groups in training. The K-correction was of questionable use within a pilot sample. Two years later, Fulkerson et al. (20) reported that five MMPI scales significantly discriminated between pilots classified as either well adjusted or poorly adjusted.

Gorney (21) utilized the MMPI and Maudsley Personality Inventory (MPI) with a sample of 38 pilots and 12 navigators of the Royal Air Force. The profiles of the aviator group differed significantly from the general population. The intercorrelations between the individual scales of the MMPI and the MPI agreed with the findings of other non-flying populations. Although the aviator scores differed significantly from the general population means, the correlations and factor loadings remained similar.

A review of the MMPI by Hedlund (22) included an evaluation of its effectiveness as a selection instrument. In a survey of 13 research studies and several review articles, Hedlund observed that methodological problems beleaguered most MMPI investigations. With regard to the MMPI as a selection device, Hedlund stated, "There is an evident scarcity of validity studies on the MMPI in selection and placement. Also, the few studies which have been conducted have found little or no relationship between any MMPI score and job performance" (22, p. 84). Similar conclusions were drawn 8 years earlier by Voas et al. (13) in an examination of the MMPI for use in naval aviation training. Comments regarding this instrument included the following: test is too long; it is not sufficiently valid; it is fakeable; and the type of attrition (pre-flight failures) that it predicts is not very costly to the Navy.

Eysenck Personality Inventory (EPI). The EPI has been used to study the relationship of social interaction style to flight training performance (23,24). The EPI, a self-report inventory that measures extraversion-introversion and neuroticism-stability, was used to predict aviation training failure (24). Jessup and Jessup (24) utilized a pass/fail criterion to predict success in training with a British Royal Air Force sample. They found that a large number of failures (60%) occurred in the neurotic-introvert quadrant. In contrast, only 14% in the stable-introvert quadrant failed flight training. Green's results (23) using the introversion-extraversion scale from the Maudsley Personality Inventory (MPI) with 80 naval aviation training candidates failed to support social interaction style as the factor responsible for prediction. Furthermore, Green found no significant differences between those individuals who voluntarily withdrew and those who had completed at least 1 year of flight training. This suggests that personal stability, rather than social interaction style, accounts for the success in prediction and warrants further investigation and cross-validation.

Personality Research Form (PRF). A recently developed personality instrument is the PRF by Jackson (25). The PRF was cited by Anastasi (26) as the test most clearly illustrating the multistage process for building validity into a test. It was used as one of a battery of tests to predict completion of U.S. Air Force navigator training (27). It significantly increased prediction beyond that accounted for by standard preselection entrance tests. Its inclusion in a model with cognitive tests increased the multiple R from .40 to .46.

Psychometrically, a personality test must possess high reliability and validity (16), not be susceptible to response bias (28), and, in terms of prediction, explain the appropriate personality dimensions and the relevant task performance (29). In a review of personality instruments (30), Kozlowski cited the PRF as the only test that satisfies all of these criteria. Kozlowski notes that the PRF is a self-report inventory based on Murray's list of psychogenic needs in which response bias is minimized. Research on the PRF demonstrates convergent and discriminant validity and high internal consistency

(25). Additional psychometric information is available in the PRF Manual (25).

The PRF has demonstrated consistency in generalizability across different populations within the Canadian Armed Forces (31). Joaquin (32) used the PRF to study undergraduate pilot training performance in the Canadian Forces. Joaquin concluded that successful trainees displayed a significantly higher degree of instrumental aggressiveness and interpersonal/leadership traits, while students who failed flight training displayed high aggressiveness scores and low interpersonal/leadership.

California Psychological Inventory (CPI). In contrast to the MMPI discussed previously, the CPI was developed to assess "normal" personalities. It was administered to 315 incoming naval aviation candidates to determine its effectiveness in predicting flight training success. Bucky and Ridley (33) found that CPI profiles of aviation candidates who complete training and those who dropped out of training at their own request, are almost identical; only the communality scale of the inventory is significantly different. They suggested that those who complete flight training are "more dependable, tactful, sincere, realistic, and conscientious,... and have more common sense and good judgment than the student who drops out of the program." However, of the 18 scales in the CPI, 1 scale would be expected to achieve significance at the .05 by chance alone. In applying the Tukey post-hoc test to the data, this difference did in fact disappear. In summary, the CPI has, in general, been of little value in predicting success in aviation training (15).

Cornell Word Form (CWF). The CWF (34) was initially developed by the Cornell University Medical School for the military during World War II. It was designed to mass-screen psychiatric problems, thus the test is short and usually requires only 5-15 min to complete. The questionnaire consists of 80 items; each item contains one stimulus word and two response-choice words. Respondents choose the word between each response pair that they associate most closely with the stimulus word. The items are highly sensitive to response bias, especially in a screening situation (35).

The CWF received some attention as a prediction instrument for aviation selection. Barry et al. (36) identified a small but significant number of aviation students who adjusted poorly to flight training based on CWF scores. Trites and Kubala (37) found a significant relationship between CWF and success as an Air Force pilot and reported significant correlations between the CWF and Personal Inventory tests. They suggested that the successful pilot is relatively free from, or tends to deny, somatic complaints or symptoms that are characteristic of maladjusted individuals.

State-Trait Anxiety Inventories and Related Scales. In a study by Green (23), the anxiety scale from the Maudsley Personality Inventory (MPI) was used to isolate potential voluntary attrites from the Navy's aviation training program in Pensacola, Florida. Those who later failed training scored significantly higher on this scale compared to those that successfully completed training.

Fleischman et al. (38) studied the relationship of five personality scales to success in naval aviation training. Student scores on two of the scales, the Taylor Manifest Anxiety Scale (TMAS) and the Alternate Manifest Anxiety Scale (AMAS), were then related to the flight training criteria of pass-fail, flight failure elimination, and voluntary withdrawal. Significant correlations were

obtained between the TMAS and pass-fail ($r = -.10$) and voluntary withdrawal ($r = -.16$). Performance on the AMAS was unrelated to the flight criteria measures.

Bucky and Spielberger (39) administered the State-Trait Anxiety Inventory (STAI) to 316 naval aviation candidates. They found that the level of anxiety at the outset of flight training was related to whether or not the student completed flight training. Students who scored high in both state (transitory anxiety or how one feels at the moment) and trait anxiety (anxiety proneness or how one generally feels) during the first week of training were most likely to attrite from the training program. Students who attrited during the early training stages tended to be higher in state anxiety during their first week of training than those who either continued or attrited at later stages of training. Those candidates who attrited as flight failures were significantly lower in both trait and state anxiety than those who attrited for other reasons. In another STAI study ($N = 8$ student pilots), Krahenbuhl et al. (40) determined that "inferior" students experience greater stress in the T-37 undergraduate pilot training program than do superior flight students.

Although these studies demonstrate that anxiety can be used as a predictor of flight training performance, another study (27) of navigation students given the STAI prior to entering Air Force flight training found no relationship between anxiety and completion of training. In summary, the STAI and other related instruments appear worthy of further attention as potential predictors of success in aviation training. The available data suggests that anxiety measures may only be useful after a student enters flight training as opposed to an entrance selection tool.

Cattell Sixteen Personality Factor (16PF). The 16PF was developed by Cattell et al. (41). According to Bartram (42), analysis of the 16PF and the EPI as predictors of passing advanced rotary wing training in the Royal Air Force ($N = 62$ aviation trainees) revealed that the 16PF was "extremely promising," but the author did not elaborate further (44). Bartram's Microcomputerized Personnel Aptitude Tester (MICROPAT) data indicated the main differences between flight successes and failures occurred on scales C, O, I, and N as predicted by Cattell et al. (41), with smaller differences on other scales. Those who passed training were more "emotionally stable" (scale C), lower in "susceptibility to anxiety and depression" (scale O), relatively "aggressive and competitive" (scale I), and "emotionally detached" (scale N). The 16PF profiles of the applicants strongly resembled those obtained from a sample of U.S. Airlines pilots and were noticeably different from the general population. Bartram suggested that candidates applying for pilot training may already be a select group. Candidates who were tested after passing standard selection procedures were not noticeably different on 16PF measures from nonpreselected samples of applicants. This indicates that for pilot selection the 16PF is relatively immune from distortion through faking. Where both EPI and 16PF data were available on the same individuals, the 16PF alone differentiated between commissioned and non-commissioned groups of applicants. Bartram's 16PF study indicates that information about personality in the Royal Air Force may increase the level of prediction obtained with measures of aptitude.

The Soviet Union also has had some success with the 16PF as a tool in aviation selection (43). Although the specific methodology is unclear and sample size is relatively low (45 "successful flight cadets" vs. 27 "less

successful"), factor C, emotional stability, reliably distinguished between successful and less successful pilot cadet groups. Successful student pilots were significantly more stable, which agrees with Bartram's results (42). Less successful pilot cadets were students who were eliminated from the flight training program for flight failure. The data are reported in such a way that it is difficult to determine the direction of any other differences between flight successes and failures with the Soviet's use of the 16PF. In fact, the instrument that is referred to as the "sixteen-factor personality inventory" may not be the instrument originally developed by Cattell (no citation is given for the instrument), and the inventory used may not be an accurate translation of the 16PF. Additionally, the investigator did not report when the cadets were tested, if they were a pre-selected sample, or if the personality factors made any unique contribution to prediction. Even the author concluded that the connection between the personality factors and flight performance was ambiguous.

Further support for the Cattell 16PF (44) as a tool for predicting success in U.S. Navy pilot training was completed as part of a larger effort (38). Factors O, N, C, and I added step increases to a multiple R of .024, .018, .005, and .003, respectively, in predicting a pass/failure flight criterion of more than 500 Navy and Marine aviation candidates. The regression analysis included current selection test variables, aviation ground school grades, and four additional personality instruments. Factors C, O, and I added unique variance to a multiple R in predicting both pass/voluntary withdrawal and pass/non-medical attrite criteria, although to a lesser extent. Point-biserial correlations between the pass/failure flight criterion and the 16PF indicated that only the O scale was significantly related ($r = .12$, $p < .01$). Factors C and I were significantly related to the pass/voluntary withdrawal criterion ($r = .13$, $p < .01$ and $r = -.09$, $p < .05$, respectively), and factor C was significantly related to the pass/non-medical attrite criterion ($r = .10$, $p < .05$). Although the individual contribution of each element of the 16PF to the prediction of the three dichotomous criteria was not available, Table 2 shows the additional variance accounted for using the personality variables (Cattell 16PF, Taylor Manifest Anxiety Scale, Alternate Manifest Anxiety Scale, Pensacola Z Scale, and the Adjective Check List) in the regression model.

TABLE 2. Multiple Point-Biserial Correlations between Predictor Variables and Three Dichotomous Criteria.

	Pass/fail	Pass/withdraw	Pass/non-medical attrite
Personality scales excluded:	.359	.150	.286
Personality scales included:	.425	.270	.381

All increases in the multiple R were significant beyond the .01 level. The results show promise and agree with other work (45).

Omnibus Personality Inventory (OPI). The OPI (46) was developed for a homogeneous population similar to military aviation students. The OPI was constructed for research on college attrition. It emphasizes intrinsic motivational factors as differentiated from extrinsic factors in learning. It is a self-administered paper-and-pencil test that consists of 385 true/false items that yield 15 scales. Because of past success with the OPI in predicting attrition from college (47) and its orientation toward attitudes in a new learning environment, it was used in an attempt to predict success in naval primary flight training (48). The authors concluded that certain OPI scales, the Theoretical Orientation (TO) and Anxiety Level (AL) subscale scores, do predict student naval aviator success in flight training beyond that accounted for by standard selection test scores. Cross-validation, however, resulted in negating the predictive validity of the OPI scores generated by the first population. The cross-validation indicated that the standard selection scores survived revalidation, but the OPI accounted for less than 0.5% of the variance with the second sample.

Group Embedded Figures Test (GEFT). Field independence successfully predicted graduation/elimination for 1199 students undergoing Navy primary flight training (49). The findings were replicated with a second sample of 1265 Navy student pilots with no decrease in statistical significance (50). In terms of simple correlation, field independence was a better predictor of pass/attrite ($r = 0.146$) than any of the four screening predictors currently in use. Because all subjects were already admitted to naval primary flight training, current aviation selection test scores and the GEFT were included in a regression analysis using a graduation/elimination criterion. Field independence was able to account for an additional 1.6% of the variance beyond that achieved with the existing screening devices. The multiple correlation between graduation/elimination and all five predictor variables was .19; if field independence is removed from the regression equation, the correlation drops off to .15. This decrease in correlation is significant beyond the .0001 level. Finally, the partial correlation between field independence and graduation/elimination controlling for the other four variables was .114. Thus, most of this relationship is indeed new information independent of the current screening variables. In general, the correlations are all quite low, which is expected as the subjects were already preselected on four of the five predictor variables used in the regression equation.

Concurrent validation for the GEFT was provided by Cullen et al. (51) using a similar instrument, the Rod and Frame Test (RFT). Both the GEFT and the RFT require disembedding a stimulus from its surrounding visual field and provide a measure of field independence/dependence (52-55). Cullen et al. (51) found that their sample of 149 commercial airline pilots was significantly more field independent than a group of aerospace engineers and college students (56). The only measure of field independence for this sample was taken after the subjects already had established careers in aviation. Thus, it is not known whether the commercial pilots in this study had high field independence scores that resulted from flight training experience (mean flight time was 2,454 h) or if field independence at the beginning of their flight careers contributed to their success as pilots.

An interesting aspect to both the GEFT and the RFT is the measurement of a personality variable that does not use a standard personality inventory item format. Both the GEFT and RFT use geometric relations as stimuli, and as such, are not as susceptible to "faking" as are most personality instruments. A substantial body of literature exists, however, that suggests that the GEFT and RFT actually measure spatial visualization skills rather than a personality trait (57-64). As discussed earlier, personality traits are relatively enduring and highly resistant to change under normal circumstances. Thus, one can assume that a personality test will yield a measure that is not continually and rapidly changing. The literature indicates, however, that scores on the GEFT shift toward field independence with test experience (52,58,65), with practice and training in spatial skills (60), or when subject groups are matched for high spatial ability (64). Considering studies that have investigated the effects of practice on the GEFT, what is actually measured appears to be a trainable spatial ability rather than a stable personality trait. As such, success in predicting graduation/elimination in flight training using the GEFT may be attributed to a relationship between spatial ability and flight performance. Furthermore, its value as a selection device is questionable if relatively minimum amounts of training or practice can substantially affect the score achieved on the instrument.

Edwards Personal Preference Scale (EPPS). The EPPS (66) measures 16 personality needs by a 244-item, forced-choice inventory derived from Murray's theory of human needs. Although the EPPS can significantly differentiate between military jet pilots and published standardized norms for males (67) and females (68) on 15 of the component subscales, it has not successfully predicted performance in primary flight training (69). Peterson et al. (69) found that the only significant difference between successful flight students and attrites was that attrites are significantly higher in need for endurance. This finding is contrary to the expected direction and is likely a chance result. Nonetheless, the EPPS consistently has generated a typical personality profile for pilots (70-73). The personality profile is a constellation of elements that are high in achievement, dominance, change, heterosexuality, exhibitionism, and aggression, and low in nurturance, nurturance, deference, abasement, order, and affiliation. The personality type attracted to aviation is adventuresome, oriented toward the demonstration of competency and achievement, and decidedly heterosexual (72). Ashman and Tefler (74) used the EPPS to compare samples of Australian Air Force pilots, trainee commercial pilots, and males drawn from the general community. Four significant effects were found for individual subscales; three (achievement, affiliation, and nurturance) correctly identified Australian Air Force fighter pilots. Commercial pilot trainees scored significantly lower than the community sample

on succorance and nurturance. The data suggest the EPPS may consist of several related personality dimensions. One of these dimensions, "sociability," successfully discriminated fighter pilots from the general community.

Strong Vocational Interest Blank (SVIB). The SVIB has demonstrated some success in the prediction of aviation training outcome. Its premise is that individuals with similar interests, needs, and qualities as those already within a specific occupational group would likely be suited for a similar occupation. The SVIB contains 325 items grouped into 7 major components. A study by Robertson (75) resulted in specific standard scales yielding almost no validity in predicting job satisfaction and little validity in predicting career motivation of naval aviators.

Guinn et al. (76) used the SVIB with a sample of Air Force cadets. Three predictor models were developed using the SVIB, Officer Biographical (OB), and Attitudinal Survey (AS). The SVIB model correctly identified 38% of all attrites but incorrectly identified only 10% of those that graduated, which equalled a 72% rate of correct classification. The OB model increased the classification rate from 65 to 68%, while the AS model improved the classification rate from 65 to 67%.

Doll et al. (77) administered the SVIB to aviation officer candidates to determine whether vocational interests of students who successfully completed naval flight training were different from those withdrawing voluntarily. Subjects who completed flight training performed significantly higher on math, science, and mechanical interest scales. In relation to the occupational scales, successful candidates scored higher on the scientific and technical scales. The authors concluded that although some overlap did exist between the SVIB and primary selection tests, the SVIB added unique variance to predicting training success. Further, because the Navy flight training program possesses a strong math-science orientation, those sharing these interests are more likely to be satisfied in Navy flight training.

Jenkins Activity Survey for Adults (JAS-C). Developed by Jenkins et al. (78), the JAS-C is a 52-item multiple-choice format questionnaire that is best known for measuring the Type A behavior pattern. The JAS-C has three subscales: Factor S (Speed and Impatience), Factor J (Job Involvement), and Factor H (Hard-Driving and Competitive).

Applying factor analysis, Spence et al. (79) derived a new measure from the JAS that consists of two moderately correlated scales labeled "achievement striving" and "impatience/irritability." Achievement Striving is positively correlated with the Work and Family Orientation Questionnaire developed by Helmreich and Spence (80). Impatience/Irritability represents an extreme sense of time urgency and a very low frustration tolerance level, which results in a tendency to react to even minor distractions with irritation. Of particular interest is that although high achievement is associated with superior performance among flight crews, it appears to have no negative health implications whatsoever (79). Conversely, high Impatience/Irritability is associated with negative health outcomes, such as sleep disturbance and fatigue, along with inferior technical flying performance.

Rotter Internal/External Locus of Control (LOC). The LOC (81) is a questionnaire containing 23 relevant items and 6 filler items in a forced-choice format of statement pairs. Scores can vary between 0 (highly

internal) and 23 (highly external). The LOC was designed to measure an individual's attributions of life events. Individuals may perceive themselves either as being in control of their behavior and life events (internally controlled) or being controlled by others (externally controlled). For example, internal scorers may believe that they are personally responsible for their safety and can take preventive steps to avoid accidents or injuries. Conversely, external scorers may believe that they have little or no personal control in accident prevention because of factors such as chance, fate, or bad luck. Rotter (81) hypothesized that people who view reinforcements as contingent on their own behavior (internals) are better adjusted than those who see reinforcements as determined by fate, chance, or powerful others (externals).

Wichman and Ball (82) administered the LOC to 82 flight instructors at a Flight Instructor Revalidation Clinic, 60 pilots at airports and flight schools, and 140 pilots at Federal Aviation Administration (FAA) safety clinics. They found that pilots were significantly more internally controlled than the general population and that self-serving biases are held by aviators. No differences between male and female pilots across all groups were found.

WHY THE FAILURES: METHODOLOGICAL PROBLEMS AND ISSUES?

Most efforts to increase the predictive validity of aviation screening systems have some inherent methodological problems. Typically, test measurement variables are related to global criterion performance measures in training such as graduation/elimination or composite flight grades. Such performance criteria, although highly useful, have several undesirable psychometric properties and may obscure the components of skilled performance or behavioral attributes associated with the selection test measure. Presumably, a given test measure may be highly predictive of a critical performance dimension during some phase or component of flight training, but the insensitivity or impracticality of the performance criterion may yield low correlations and a consequent dismissal of the test's predictive power. Helmreich et al. (83) further point out that different combinations of predictors relate to quite different measures of performance at different points in time.

Previous studies of the use of personality indices characteristically have been piecemeal and have examined only one or a few tests related to a given overall flight performance criteria, usually a composite measure at the conclusion of initial flight training. Additionally, the vast majority of investigations used subjects that already were preselected by standard selection measures. Thus, in many cases, only simple relationships between a personality measure and a singular criterion are presented. Relatively few studies contain multiple regression models of the initial candidate selection variables. Whether a particular personality variable actually adds unique variance to predicting training success beyond the initial selection measures is not yet known. Unfortunately, efforts to relate specific predictors to reliable subcomponents of overall flight grades in primary training proved unsuccessful (84,85). The authors (84) concluded that reliable clusters of performance criteria were not embedded in the overall cumulative flight grade. This was attributed to a wide disagreement among instructor pilots as to which individual measures of flight performance were used most in evaluating differences in student performance.

Research to develop subcriteria embedded in the more global criterion of graduation/elimination met with similar failure. The Army Air Forces Pilot Project (86) attempted to develop subcriteria against which specific selection measures of aptitude could be validated. Restricted range in grading flight performance was identified as a major reason for the lack of success (84). Flight students were graded subjectively in one of several categories, "A-F," with the majority receiving a "C." This was due to the emphasis on determining which students would not successfully complete flight training as opposed to providing a normal distribution of grades to differentiate among students who were successful.

All subjective evaluation systems have inherent problems that affect the validation of selection devices and personality measures. Subjective differences in grading standards introduce a source of error variance that is unrelated to a student's flying ability. Initial work by the Army Air Force (8) revealed enormous differences between check pilots (pilots that evaluate other pilots both in flight and in simulators) both within and between the various training commands. Even with a global measure of training success, differences in attrition rates ranged from 10 to 60%, which makes the accuracy of graduate-versus-eliminee categories questionable measures of student performance.

The halo effect phenomenon (86) is related to the restricted range problem in the flight criteria. Typically, check pilots and instructor pilots consider a student's past performance when preparing a current evaluation. Correlations between performance measures for different maneuvers and procedures tend to be high, suggesting the presence of a strong halo effect. Grading tendencies of flight instructors to the average or "norm" can also reflect a de-emphasis towards comparing successful students during primary training. Current military primary flight training systems also require instructor pilots to provide a written explanation when an assigned grade is other than average. In other words, instructors who assign grades that are not average are required to provide additional time-consuming documentation. A related issue is the reliability of assigned flight grades. The importance of this methodological concern to pilot selection was noted over 40 years ago (8). Studies conducted during the Army Air Force Pilot Project indicated that landing performance measures correlated near zero for repeated measurements on the same maneuvers during different days using different aircraft and instructors with the same students.

The candidate population itself poses a methodological problem in validating personality instruments. Most personality inventories and clinical diagnostic tools were developed for testing heterogeneous groups. Military aviation candidate populations tend to be comparatively homogeneous. Typical entrance requirements include a 4-year college degree, rigid medical requirements, and initial aviation screening tests. Application to a flight training program in itself reflects a general interest in aviation. Most applicants are males in their early twenties as well (military age standards partially account for the similar age factor found in the candidate population). All of these factors combine to result in a rather unique, homogenous population that severely restricts the sample population at the outset.

Another reason for the few personality measures that discriminate at the selection level may be that no personality differences exist. This is plausible, given that application to a military flight training program is

completely voluntary and that military aviation attracts a particular personality profile or type. An alternative possibility is that present personality tests are not sensitive enough to detect the existing differences.

Others maintain that personality measures can only effectively predict actual job performance and not training performance. Helmreich (87) emphasizes that "deficiencies in the criterion lead to overemphasis on some predictors and the neglect of others." Helmreich et al. (83) reported a link between personality and performance, called the "honeymoon effect" of motivation on performance. They believed that the honeymoon effect was the maximum effort that many job prospects exert in order to obtain a coveted position or job. Only after the "honeymoon" period ends, do the underlying personality dispositions become significant determinants of behavior. Their study suggests a major weakness in using initial training performance as the selection criterion. In the same study, personality and motivational factors measured prior to employment proved to be good predictors of job performance. This prediction was obtained only after the subjects had been out of training and on the job for more than 3 months. The predictors were unrelated to performance both in training and after initial release to the workforce.

Helmreich (87) administered the Extended Personal Attributes Questionnaire (EPAQ (88)) and the Work and Family Orientation Questionnaires (WFOQ (80)) to a group of civilian pilots. The EPAQ measures positive and negative clusters of instrumental and expressive traits; the WFOQ evaluates three aspects of achievement motivation and interpersonal competitiveness. These personality measures were compared to ratings by check pilots. The results indicated that the trait constellations of instrumentality and expressiveness, along with components of achievement motivation, were significantly related to this operational criterion. The better pilots scored higher on instrumentality, expressivity, and high mastery needs, while poorer pilots scored higher on aggressiveness.

Test response bias is the inability to obtain a true measure of an individual's character, which is usually attributed to response sets. It is often cited as responsible for the lack of validity in predicting a flight training criterion (6,11). Social desirability or "faking" is the response set or attitude that has received the greatest attention. As Anastasi (12) pointed out, respondents can easily detect the most socially desirable or acceptable response choices in the majority of personality inventories. In military aviation testing scenarios, candidates generally will respond to create the most impressive image of themselves or to their perception of the "aviator personality." These circumstances provide very little variance on personality measures between respondents (11). Acquiescence, or the tendency to respond in a consistent but inaccurate fashion, is an additional response set that can affect the predictive validity of an instrument. Many personality inventories are structured such that all "true," "yes," "a," et cetera responses are keyed positively for the personality dimension of interest. This type of format is susceptible to some respondents answering "yes," "no," or "middle-of-the-road" for all questionnaire items. This type of response pattern does not accurately reflect the trait being measured.

Commercial availability of personality instruments is a final consideration that is often overlooked in personnel selection. Assuming that a personality test does meet the aforementioned criteria, its predictive value will probably

decline steadily within a short period of time, which is common with all measurement devices. Nonetheless, the commercial availability of personality instruments compromises test validity and provides an impetus for accelerated deterioration. This is true especially when the "score" on the instrument may determine acceptance or rejection into a military flight training program. We already know that job candidates fake personality inventories to gain employment (89,90). Within 2 years, preparation and "coaching" for the instrument may be found in commercially available guides (i.e., Officer Candidate Tests) (91), and the test could be compromised.

Considering these disadvantages, we recommend investigations of non-inventory techniques and methods of measuring personality that might provide useful additional predictions of aptitude measures. One such approach could be toward the development of measures in which the personality dimension of interest is "masked" or concealed from the candidate.

EMERGENCE OF AUTOMATED BEHAVIOR-BASED INVENTORIES

The need to improve the selection of military aviation applicants, along with recent advances and innovations in computer technology and psychological theory/measurement (26), have stimulated interest in computerized assessment. This new emphasis is partly responsible for the use of performance tasks, rather than paper-and-pencil tests, to avoid verbal and cultural biases. In the past decade, several computer-based experimental aviation selection test batteries have evolved, along with an interest in reaction and response-time measures as dependent variables. In a recent review, Bartram and Bayliss (92) argued that while the automation of existing paper-and-pencil tests has some marginal advantages (and some disadvantages), the real future of automated testing is in the development of (a) new tests, particularly new types of tests; (b) adaptive and tailored testing techniques; and (c) rule-based item-generation by computers. The following discussion presents some of the innovative approaches to personality assessment using computer-based systems.

ENGLAND (ROYAL NAVY)

The Microcomputerized Personnel Aptitude Tester (MICROPAT) was developed for the British Army Air Corps. The current MICROPAT contains two main categories of tests--psychomotor ability and information management ability. The latter category involves a greater cognitive element, which includes tests of risk taking (RISK), scheduling ability (SCHEDULE and LANDING), time-sharing (DUALTASK), and decision making (SIGNAL and PLANE). The RISK test is the only instrument designed specifically for personality assessment.

Bartram reports an evaluation of the MICROPAT RISK task based on 53 subjects (27 males and 26 females). The risk task consists of two conditions (A and B). Four blocks of 20 trials each are administered using an A-B-B-A design. Subjects are instructed that "important documents" have been left at eight locations and that they must send out a team of men to collect the information. The problem, they are told, is that one of the locations is set up for an ambush by the enemy. If the team is sent to the ambush location, they will be caught and sent back without the documents. Each document is worth 10 points, therefore a maximum of 70 points can be obtained for each trial. The ambush is randomly programmed prior to each trial. The subject is instructed to get as high a total score as possible for each of the trials. In condition A, an ambush is

set on every trial; whereas in condition B, an ambush is set on only half the trials. Primary measures of risk include: 1) mean number of keys pressed per trial, 2) mean number of keys pressed for condition A (blocks 1 and 4), and 3) mean number of keys pressed for condition B (blocks 2 and 3). Bartram (42) reports high internal consistencies for primary measures 1 and 2 and an increase in riskiness (number of keys pressed) with practice. In addition, Bartram reports sex differences; males adopted a more risky strategy than females. Information about the use of the risk task to predict performance in flight training is not yet available, although Bartram maintains it is under investigation.

U.S. AIR FORCE BASIC ATTRIBUTES TESTS (BAT)

In 1981, the United States Air Force began a large-scale effort to determine the validity of a computer-based test battery for pilot selection and classification. Known as the Basic Attributes Tests system or 'BAT' (93), the BAT consisted of 15 component tests at its inception. Although the primary emphasis of the BAT was directed toward measuring psychomotor, cognitive, and perceptual skills, six tests were included to measure personality and attitudinal characteristics. Personality tests that were included or developed were: the Dot Estimation Task, Risk-Taking, Embedded Figures, Self-Crediting Word Knowledge, Activities Interest Inventory, and Automated Aircrew Personality Profiler.

Dot Estimation Task

The Dot Estimation Task was a paper-and-pencil test developed by the Air Force in the early 1960s (94) to measure compulsiveness/decisiveness. Subjects view simultaneously two boxes containing an arbitrary number of dots; one of the boxes has one more dot than the other. The subject is instructed to determine which of the two boxes contains the greater number of dots but is not explicitly told to count the dots. The task has a time limit of 5 min with a maximum of 55 box pairs. Compulsiveness/decisiveness is determined by the number of pairs the subject attempts in the time allotted. As a computerized measure, reaction time for each response is also possible, but reliability and construct validity have never been established for this measure. Results (95) indicate that the instrument has little if any predictive validity to either a graduation/elimination criterion in undergraduate pilot training or instructor pilot recommendation for a follow-on training assignment (fighter or non-fighter aircraft).

Risk-Taking

Ten boxes are presented in 2 rows of 5 each for a total of 30 trials. The subject is told that 9 of the 10 boxes contain a reward (points), while the remaining box is a "penalty box." If a selected box contains a reward, the subject is allowed to keep it, however, if a penalty box is selected, the accumulated points for that trial are forfeited. Twelve of the trials have no penalty box, and the subject is not aware of this deviation in the task. The average number of boxes chosen provides a measure of risk-taking tendencies. Subject response time and number of boxes chosen for both the "risk" (penalty box present) and "no-risk" (penalty box absent) conditions are recorded.

Embedded Figures Test

The Embedded Figures Test is a computerized version of the original paper-and-pencil test developed by Witkin (52). Some modifications to the original version were necessary for mass implementation on a computer screen. For each trial, the subject is presented with a simple geometric figure and two complex figures and instructed to indicate which of the two complex figures has the simpler figure embedded within it. The test was included in the BAT system to assess the factor of field dependence/independence. This version of the task has 30 trials. Reaction time and accuracy are the measures of interest. Prior research has shown that the Embedded Figures Test has some predictive utility and warrants further consideration (50). As discussed earlier, however, any predictive power is probably due to a strong spatial component. A U.S. Air Force study using 1,977 pilot training candidates suggests that performance on the BAT Embedded Figures Test is not related statistically to flying training performance (96).

Self-Crediting Word Knowledge Test

The Self-Crediting Word Knowledge Test, an instrument to measure self-confidence, requires the subject to choose the closest synonym to a target word from five responses. The task is essentially a vocabulary test of 30 trials in which the target words become increasingly difficult. Before each set of 10 trials, subjects are instructed to make a "bet" that reflects how well they expect to do, with the understanding that the task becomes increasingly difficult. The average number of points bet (or "risked"), reaction time for correct responses, and percentage correct are recorded for each subject. Subjects who are more cautious (bet less and take longer to respond) are more likely to complete training successfully (95).

Activities Interest Inventory

The Activities Interest Inventory is a questionnaire designed by the U.S. Air Force to sample an aviation candidate's interests in a variety of activities. The subject is presented with 81 pairs of activities that differ in risk and threat to physical harm. For each activity pair, subjects choose a response based on the assumption that they have the necessary ability to perform each activity. The number of high-risk options chosen and the average response time for each activity pair are the principal measures of interest.

Automated Aircrew Personality Profiler

The Automated Aircrew Personality Profiler is a 202-item questionnaire designed by the School of Aerospace Medicine at Brooks Air Force Base to measure general attitudes and interests. The questionnaire is a forced-choice personality inventory with two alternatives for each item. The respondents are instructed to give the first answer that comes to mind and to respond as quickly as possible. Performance on this test demonstrates only weak validity against flying training criteria (95).

Recently, Siem et al. (95) evaluated five of the BAT personality instruments. Data on the Automated Aircrew Personality Profiler were not available. The personality tests were administered to 883 Air Force pilot candidates to assess their utility in predicting training outcome

(pass/fail) and advanced training recommendation (fighter or non-fighter aircraft). Both criteria were treated as dichotomous variables. Acceptable reliabilities were reported for all five measures for use as selection instruments.

No single test or individual dependent measure displayed a consistent pattern of validity to both criterion measures. The test for self-confidence (Self-crediting Word Knowledge) appeared to be the only instrument that contributed to predicting successful completion of flight training, with successful candidates demonstrating more caution. The only dependent measure that exceeded a correlation of .10 with the pass/fail criterion was the correct response reaction time for the Self-crediting Word Knowledge task ($r = .12$, $p < .001$). The multiple correlation for the Self-crediting Word Knowledge test was .14. No measure displayed a significant relation to instructor pilot recommendation. Although significant differences were not observed, data comparing 259 attrites with 488 successful graduates indicated a general trend toward cautious responding by students who completed training. These candidates chose fewer high-risk items on the Activities Interest Inventory, required more time and completed fewer trials on the Dot Estimation Test, and had higher percentage correct scores for the Dot Estimation Test. These findings, taken in conjunction with the results of the Self-crediting Word Knowledge task, were interpreted as a more cautious decision-making style on the part of successful candidates. This interpretation, however, was not supported by results from the Risk-taking task, which was intended to measure risk tendencies in decision making.

In summary, personality variables analyzed by the Air Force show very little promise for use in selecting or classifying aviation candidates. Further work is ongoing at the Air Force Human Resources Laboratory in San Antonio, Texas, to determine if the Self-crediting Word Knowledge Task adds unique variance to the current prediction model, even though only a weak relationship exists between the instrument and the graduation/elimination criteria. Additional research efforts are focused on improving the existing Self-crediting Word Knowledge Test and evaluating the test's construct validity. To assess specifically what the test is measuring, more traditional personality tests of characteristics, such as self-confidence (88), are being administered to Air Force flight personnel with varying levels of experience.

U.S. NAVY PERFORMANCE-BASED PERSONALITY TESTS

Dot Estimation Test

The U.S. Air Force Human Resources Laboratory (94) attempted to circumvent the problem of response bias on personality devices by developing a task in which the personality trait of interest was masked. The major difference between the Navy and the Air Force versions is that the Navy test has 50 presentations and takes 6 min. compared to 55 presentations in 5 min for the Air Force Task. As previously stated, the task was developed to provide a measure of the trait compulsivity-versus-decisiveness, assuming that the compulsive individual will require more time in making a choice as a result of vacillation between two alternate choices. Another assumption is that "re-checking" behavior, a well-documented component of compulsivity, will provide a good measure of compulsivity in general.

The Air Force results indicated that the Dot Task is not a valid predictor of either pass/fail in primary flight training or instructor recommendation for jet aircraft. In a recent Navy study, Gibb and Dolgin (97) also found no significant differences between training success and attrite groups in relation to flight grades or pass/fail. To estimate task reliability and construct validity, the Dot Task was administered with either of two paper- and-pencil compulsivity instruments to 153 college students (98). Four weeks later, 90 subjects were retested on the Dot Task and the alternate compulsivity instrument. The Dot Task had no relationship to previously validated compulsivity measures, and it lacked construct validity in its present form. The task was found to have a modest test-retest reliability of .64. Comparatively lower test-retest reliabilities could be expected with non verbal behaviorally based measures than with traditional paper-and-pencil measures, mainly because responses to paper-and-pencil measures can be remembered during retesting and cause subjects to respond consistently across testing sessions. Because nonverbal measures lack this information base, they tend to demonstrate deflated reliabilities. Possibly, construct validity could not be established for the Dot Task because of two inherent flaws in the presentation and instructions for the task. First, the instructions clearly informed subjects that the task was 6 min long and that they were to respond as quickly and accurately as possible to as many of the 50 pairs of field comparisons as they could. Imposing a time constraint on the task may have suppressed the compulsive trait of rechecking, which the task was intended to measure. Secondly, the task provided little personal consequence (reward or penalty) related to accurate or inaccurate responding; individuals may only exhibit those behavior patterns in personally relevant areas of life. In summary, although the Dot Estimation Task has not been validated, it does represent an attempt to tap personality dimensions using a masked technique to overcome problems of response bias.

Risk Taking

Long and Shelnutt (99) reviewed risk-taking theory and research from its antecedents in economic theory of the 1950s to the role of risk-taking in decision making in the 1970s. Their conclusions about risk taking measures were much the same as other writers (100), that is, that numerous and varied measures were purported throughout the decades to assess "risk." Specifically, risk measures encompassed diverse behaviors, such as goal setting and betting preference (101,102); skillplay, such as ring-tossing and shooting (103, 104); and opinion questionnaires (105).

Risk-taking tendency is a primary component of decision making, which is widely cited as critical to piloting (106). A number of tasks exist that purport to measure an individual's risk-taking tendencies, including the risk-taking task (106), the sequential gamble (107), and the choice dilemma instrument (108). According to the portfolio theory (Coombs study cited in 106), individuals have a stable level of risk to which they are willing to engage. The level of risk is typically measured by the individual's willingness to accept a given level of probability to obtain a payoff and by the decision response time or latency. Because piloting decisions are often made under time constraints, response times of risk-taking behavior are important to measure.

Shull et al. (109) conducted an initial validation of the Navy test for measuring risk-taking tendencies in 440 student naval aviators. The Navy risk test is essentially a computer-based gambling task consisting of 3

sessions with 10 trials in each session. For each trial, the subject is presented with a matrix of squares identified by numbers. At the beginning of each trial, one square is a penalty square, which causes a loss of points, and nine are reward squares. During session 2, two randomly selected penalty squares (for each trial) provide an opportunity to assess changes in response strategy to a more "risky" situation. The subject is allowed to select any of the squares, one at a time, and if the selected squares contain a payoff (points), the subject may keep it. Measures indicating increased risk-taking consist of increases in number of responses made (squares selected) and decreased response latency in making those selections. Results from the risk test were compared to students' raw scores on the navy's primary flight candidate selection battery and actual grades from flight training. The number of squares selected during the first session and the pass/attrite criteria were significantly correlated, which indicated that increased risk-taking is associated with completing primary flight training. The authors also found significant correlations between this particular measure and both the aviation indoctrination and cumulative flight grade scores, although in a direction indicating that decreased risk-taking is associated with higher grades in these areas. If present results are any indication, this test or some revised version of it may hold promise as an effective pilot candidate screening device. However, in a U.S. Air Force study Siem et al. (95) found no relationship between risk taking behavior and pass/fail outcome with a sample of 883 pilot candidates.

SCANDANAVIAN FORCES

Defense Mechanism Test (DMT)

The DMT was devised in 1961 in Sweden (110). Since then, it has undergone continuous development and wide application in personnel selection, notably pilot selection, in Europe (111-114). The test is based on three basic theoretical principles: the theory of projective techniques, the concept of percept genesis (PG), and the psychoanalytic theory of defense mechanisms. In projective techniques, a subject is presented with a situation (e.g., a picture) in which objective cues are minimized to effect considerable ambiguity in the content of the external stimulus. With respect to the DMT, subjects view pictures containing a central figure or hero with whom they are supposed to identify and a threatening peripheral figure. The DMT is a projective personality test in which a picture displaying psychologically threatening aspects is shown repeatedly to a subject under conditions of increasing exposure times ranging from 10 to 1000 ms. At the shorter exposure times, only a partial perception of the picture is possible. Vulnerability to perceiving threats is measured by comparing the subject's responses to the same pictures at longer exposure times. The premise is that a subject who "sees" the threat early will spend less psychological energy restructuring the world and, therefore, can identify and handle difficult situations better than a person who is unwilling to deal with the world as it really is. Production and maintenance of defense mechanisms require considerable energy, which leaves fewer resources available for coping with stress present in occupations such as flying and deep-water diving. Most people have defense mechanisms, but as the amount of defensive organization increases, the ability to cope with external stress decreases.

The PG concept maintains that perception is not an instantaneous function; it is a process that develops over time. During the development

of a percept, before the representation of the external stimulus becomes clear in consciousness, this developing representation is vulnerable to modification by the needs and motives of the perceiver, that is, aspects of the personality. In situations such as those used in projective tests where the objective stimulus is ambiguous, such "distortion" of perception has a greater likelihood of taking effect, and analysis of the early stages in perception of such stimuli is assumed to yield information regarding the individual's personality.

Finally, psychoanalytic theory of defense mechanisms (see 115 for review), as applied to PG, states that certain classes of stimuli are recognized very early in the perception process as being "dangerous" or "threatening" to the individual's ego, representation of self, and "psychological" security. The salient point is that these stimuli evoke reactions designed to protect the ego from the threat, that is, ego defenses or defense mechanisms.

The rationale for the predictive usefulness of the DMT is that the production and maintenance of defense mechanisms require considerable amounts of psychological energy. Thus, fewer resources are available to cope with stresses present in occupations such as flying. In addition, empirical data show that frequent use of certain specific defense mechanisms, such as reaction formation, tend to be associated with certain pilot behaviors. For example, accidents resulting from pilot-error are related hypothetically to an overuse of the reaction formation defense mechanism.

In the Swedish Air Force, Neuman conducted two validation studies from 1967 to 1970 and from 1975 to 1978. The criterion was inadequate adaptation to military flying (failure in basic or advanced flight training, adjustment difficulties, psychosomatic problems, flight neuroses, and flight accidents). In the first study, 31% of pilots with "poor" DMT scores were lost to the service over the 3-year period, compared to 10% of pilots with "good" scores. The accident data showed that 14% of pilots with poor scores became involved in accidents, whereas only 1% of those with good scores did. Of 14 pilots involved in flight accidents over the 3-year period, 13 would have been identified by their test scores. The second study showed that, when revised scoring weights were applied, 7% of poor scorers were classified as adapted, compared to 56% of good scorers. No separate accident data were reported. The test became a functional part of the Swedish Air Force pilot selection procedure in 1970. The Danish Air Force introduced the DMT in 1975 using methods of administration and scoring identical to those used by the Swedish Air Force. Danish Air Force results showed that 87% of poor scorers failed basic flight training as compared to 31% of good scorers.

The DMT is the last stage in a sequential selection procedure. Aviation candidates are eliminated for medical, motivational, and aptitudinal reasons, and only those remaining are administered the DMT. Based on DMT results, the rejection rate is approximately 25%. The reader should note that DMT results are not evaluated in isolation. The psychologist who administers the DMT is a member of the full selection board and has access to all other information on the candidate. The psychologist's recommendation is the primary factor in the final acceptance or rejection of a candidate.

The British Royal Air Force (116) attempted to modify the test for group administration but was unsuccessful, and no conclusions as to its construct validity in the revised format could be drawn. Group administration is beset

with many difficulties that have not yet been resolved; vital responses may not be forthcoming unless elicited, distances of candidates from the projection screen vary, lighting levels may vary, and there may be interference or social support effects.

In summary, positive results with the DMT are limited to the Swedish and Danish Air Forces studies. Their data clearly demonstrate that, when used in those contexts and in the approved manner, the DMT predicts both training outcomes and flight safety criteria with a high degree of validity. The Royal Air Force experience shows that circumventing the established procedure may result in failure and inconclusive results.

THE ROLE OF PERSONALITY IN AVIATOR SAFETY

The 1980s reflected a renewed interest in personality as it relates to aviation safety using tests other than the DMT. Typically, research has been directed toward identifying the "accident prone" aviator. However, "accident proneness" is not a stable characteristic and is situationally based (117,118). Measurement of the tendency to be accident prone or susceptible would thus be difficult because the tendency varies with time. Increased risk-taking tendencies that result in mishaps would only emerge as a result of situational circumstances in conjunction with an inability to cope with increased stress levels. Alkov et al. (117) suggest that inadequate techniques for coping with stress, rather than cumulative life stress, account for the increased levels of accident susceptibility. Recent data (117,119) that compare pilots who were causally involved in mishaps with aviators involved in mishaps with no culpability suggest that pilots who made errors resulting in mishaps were poorer leaders, were less mature and stable, had undergone a recent lifestyle change, and were experiencing problems with interpersonal relationships. Alkov et al. (117) conclude that aircraft mishaps may be attributable to the non-introspective personality, but the data are post-hoc and are not based on a prediction model. Aviators involved in aircraft accidents were evaluated on numerous dimensions by accident investigation board members and through interviews with superiors, peers, and family. Information provided by the respondents was biased by the aviator having been involved in a mishap. Using personality devices to predict which individuals would be involved in future aircraft accidents would be difficult and require enormous sample sizes due to the relatively low incidence of mishaps.

Jensen and Benel (120) reviewed literature containing aviation accident data from 1970 through 1974. Their conclusions were: 1) Erroneous pilot decision-making was a factor in 35% of all non-fatal aviation accidents, and 2) faulty decision-making played a definite role in 52% of fatal mishaps. The authors noted that research on pilot judgment was sparse and, for the most part, unsystematic. They maintain that pilot judgment is trainable and can be objectively evaluated. In conclusion, they speculate that faulty judgment might result from a pilot's proclivity to situational influences such as peer reactions, fear of failure, censure from superiors or family members.

More recently, Lester and Bombaci (121) examined the construct validity of five "hazardous thought patterns," hypothesized to mediate pilot judgment. The hazardous thought pattern concept is the result of an investigation carried out by the FAA and Embry-Riddle Aeronautical University (ERAU). In response to the Jensen and Benel study (120), ERAU investigators sought to isolate the specific

thought patterns that might serve as the precursors to faulty pilot judgment. Based on a literature review and consultation with experts in the behavioral sciences, five hazardous thought patterns were identified: anti-authority, impulsivity, invulnerability, macho, and external control or resignation. A 10-item self-assessment inventory was designed to assess the hazardous thought patterns concept. Evaluating a sample of 35 civilian pilots, Lester and Bombaci (121) observed a significant relationship between hazardous thought patterns and scores on both the 16PF integration/self-concept control scale and the Rotter LOC scale. They recommended that additional research examine the way in which situational influences interact with pilot personality. Table 3 contains a description of the five hazardous thought patterns.

TABLE 3. The Five Hazardous Thoughts.*

1. Anti-Authority: "Don't tell me!"	This thought is found in people who do not like anyone telling them what to do. They think, "Don't tell me!" In a sense, they are saying "No one can tell me what to do." The person who thinks, "Don't tell me," may either be resentful of having someone tell him or her what to do or may just regard rules, regulations, and procedures as silly or unnecessary. However, it is always your prerogative to question authority if you feel it is in error.
2. Impulsivity: "Do something--quickly!"	This is the thought pattern of people who frequently feel the need to do something, anything, immediately. They do not stop to think about what they are about to do; they do not select the best alternative--they do the first thing that comes to mind.
3. Invulnerability: "It won't happen to me."	Many people feel that accidents happen to others but never to them. They know accidents can happen, and they know that anyone can be affected; but they never really feel or believe that they will be the involved. Pilots who think this way are more likely to take chances and run unwise risks, thinking all the time, "It won't happen to me!"
4. Macho: "I can do it."	People who are always trying to prove that they are better than anyone else think, "I can do it." They "prove" themselves by taking risks and by trying to impress others. While this pattern is thought to be a male characteristic, women are equally susceptible.
5. Resignation: "What's the use?"	People who think, "What's the use?" do not see themselves as making a great deal of difference in what happens to them. When things go well, they think, "That's good luck." When things go badly, they attribute it to bad luck or feel that someone is "out to get them." They leave the action to others--for better or worse. Sometimes such individuals will even go along with unreasonable requests just to be a "nice guy."

* Description of the five hazardous thought patterns. (From Human Factors, 1984, Vol. 26, p. 568. Copyright 1984 by the Human Factors Society, Inc. and reproduced by permission.)

CONCLUSIONS AND RECOMMENDATIONS

The development and application of personality tests present unique opportunities, as well as special difficulties, that might not be encountered with aptitude testing. For example, test faking and malingering are more problematic in personality assessments. As we have described, attempts to improve personality assessment have included computerization, the development of verification and correction scales, keying certain items against specific criteria, masking the dimension of interest, and the application of factor analysis to isolate more specific trait categories. Of these, computer administration and concealing the personality trait of interest appear to hold the most promise for the future of personality testing in aviation selection.

One of our main goals was to identify specific tests that warrant further research as potential prediction instruments. The majority of personality instruments reviewed were not useful for pilot selection. In some cases, methodological difficulties may have obviated more promising results. Based on the review of past and present instruments utilized in the selection of pilots, we recommend the following seven tests for continued research because they appear to be both effective in pilot selection and psychometrically sound:

1. One test that we recommend is the Defense Mechanism Test (DMT) because of its effectiveness in predicting pilot training success and its proven safety in the Swedish and Danish forces (111). The DMT is a projective personality test that has been used operationally in Scandinavian countries for the past decade. The concept of the DMT in predicting success in flight training is that the use of certain defense mechanisms may limit the amount of "psychological" energy available for handling external stress. Because the military flight training environment is highly stressful, a flight candidate with intense defenses might not immediately recognize a dangerous situation. Although the DMT is designed for individual administration and requires 1.5 to 2 h testing time, previous success with the instrument warrants further study. In addition, computerization of the DMT is highly recommended in order to identify the stimuli that are producing the effect, increase objectivity, and shorten test-taking time.

2. The Personality Research Form (25) is recommended due to its psychometric construction (26) and promising research results in the Canadian Armed Forces (31,32) and the U.S. Air Force (27).

3. The Cattell 16PF (41) has been used successfully (38,42,44) to predict success in flight training. Lester and Bombaci (121) found a significant relationship between "hazardous thought patterns" and 16PF scores. As a result of these studies, the 16PF stands out as a personality instrument requiring further investigation.

4. Another test that has achieved some success in defining pilots is the Locus of Control (81). The Locus of Control is a brief questionnaire consisting of 23 items and is easily automated for computer administration. Findings from studies (82,121) determined that pilots are significantly more internally controlled than the general U.S. population.

5. Developed by Spence et al. (79), the Work and Family Orientation Questionnaire (WOFO) has been related successfully to pilot performance (83).

The WOFO operationalizes achievement motivation into components of mastery needs, desire to undertake new and demanding tasks, work orientation, satisfaction with hard work and task completion, competitiveness, and concern with outperforming others in interpersonal situations.

6. Another recommended instrument is the Extended Personality Attributes Questionnaire (EPAQ: 80,88). The EPAQ has typically been employed in research concurrently with the WOFO.

7. The Strong Vocational Inventory Blank (SVIB: 76,77) has demonstrated validity as a predictor of success in both the Air Force and the Navy. The SVIB measures vocational interest patterns based on various preferences.

In the future, aviation selection will most likely utilize prediction of performance beyond initial training. The areas of pilot judgment, aviation safety, cockpit crew coordination, and operational flight performance interact closely with individual differences in personality, and most likely, research endeavors will be initiated toward assessing those relationships. Personality assessment in predicting training success, however, will undoubtedly receive the greatest attention as a result of the variance unaccounted for with aptitude measures and the driving force of upwardly spiraling training costs.

REFERENCES

1. Naval Education and Training Program Management Support Activity letter 7000/00B26 of 8 October 1987.
2. Damos, D. and Gibb, G.D., Development of a Computer-based Naval Aviation Selection Test Battery, NAMRL-1319, Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, FL, August 1986.
3. American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders III, Washington DC, 1980.
4. McFarland, R.A., Human Factors in Air Transportation, McGraw-Hill, New York, NY, 1953.
5. Boring, E.G., A History of Experimental Psychology, Prentice-Hall, Englewood-Cliffs, NJ, 1950.
6. North, R.A. and Griffin, G.R., Aviator Selection 1919-1977, NAMRL SR77-2, Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, FL, October 1977.
7. Fiske, D.W., "Validation of Naval Aviation Cadet Selection Tests Against Training Criteria." Journal of Applied Psychology, Vol. 31, pp. 601-614, 1947.
8. Miller, N.E., Psychological Research on Pilot Training, Army Air Forces Aviation Psychology Program Research Report No. 8., U.S. Government Printing Office, Washington DC, 1947.
9. Youngling, E.W., Levine, S.H., Mocharnuk, J.B., and Weston, L.M., Feasibility Study to Predict Combat Effectiveness for Selected Military Roles: Fighter Pilot Effectiveness, TR-MDC E1634, McDonnell Douglas Astronautics Co., East St. Louis, MO, 1977.
10. Guilford, J.P., Printed Classification Tests, Army Air Forces Aviation Psychology Program Research Report No. 5., U.S. Government Printing Office, Washington DC, 1947.
11. Ellis, A. and Conrad, H.S. "The Validity of Personality Inventories in Military Practice." Psychological Bulletin, Vol. 45, pp. 385- 427, 1948.
12. Sells, S.B., "Further Developments on Adaptability Screening of Flying Personnel." Aviation Medicine, Vol. 26, pp. 35-45, 1956.
13. Voas, R.B., Bair, J.T., and Ambler, R.K., Validity of Personality Inventories in the Naval Aviation Selection Program, NSAM-13, Naval School of Aviation Medicine, Pensacola, FL, November 1957.
14. Anastasi, A., Psychological Testing, (5th ed.), Macmillan, New York, NY, 1982.

15. Griffin, R.G. and Mosko, J.D., Naval Aviation Attrition 1950-1976: Implications for the Development of Future Research and Evaluation, NAMRL-1237, Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, FL, 1977.
16. Anastasi, A., Psychological Testing (4th ed.), MacMillan, New York, NY, 1976.
17. Hathaway, S.R. and McKinley, J.C., "A Multiphasic Personality Schedule (Minnesota): I. Construction of the Schedule." Journal of Psychology, Vol. 10, pp. 249-254, 1940.
18. Melton, R.S., "Studies in the Evaluation of the Personality Characteristics of Successful Naval Aviators." Journal of Aviation Medicine, Vol. 25, pp. 600-604, 1954.
19. Fulkerson, S.C., Adaptability Screening of Flying Personnel: Development of a Preliminary Screening Battery, USAF SAM 56-84, School of Aviation Medicine, Randolph AFB, San Antonio, TX, August 1956.
20. Fulkerson, S.C., Freud, S.L., and Raynor, G.H., "The Use of the MMPI in the Psychological Evaluation of Pilots." Journal of Aviation Medicine, Vol. 29, pp. 122-129, 1958.
21. Goorney, A.B., "MPI and MMPI Scores, Correlations and Analysis for a Military Aircrew Population." British Journal of Social and Clinical Psychology, Vol. 9, pp. 164-170, 1970.
22. Hedlund, D.F., "A Review of the MMPI in Industry." Psychological Reports, Vol. 23, pp. 507-510, 1965.
23. Green, L.R., An Exploratory Investigation of the Relationship between Four Personality Measures and Voluntary Resignation from Aviation Training, SAM 872, U.S. Navy School of Aviation Medicine, Naval Air Station, Pensacola, FL, 1963.
24. Jessup, G. and Jessup, H., "Validity of the Eysenck Personality in Pilot Selection." Occupational Psychology, Vol. 45, pp. 111-123, 1971.
25. Jackson, D.N., Personality Research Form Manual, Research Psychologists Press, Goshan, NY, 1974.
26. Anastasi, A., "Evolving Concepts of Test Validation." Annual Review of Psychology, Vol. 37, pp. 1-15, 1986.
27. Valentine, L.D., Navigator-observer Selection Research, AFHRL-TR-77- 36, Air Force Human Resources, Brooks Air Force Base, TX, May 1977.
28. Wiggins, J.S., Personality and Prediction: Principles of Personality Assessment, Addison Wesley, Reading MA, 1973.
29. Guion, R.M. and Gottier, R.F, "Validity of Personality Measures in Personnel Selection." Personnel Psychology, Vol. 18, pp. 135-164, 1965.

30. Kozlowski, S.W.J., The Validity of Personality Inventories for the Selection of Personnel: A Review of the Literature and Recommendations for Research, State Civil Service Commission, Bureau of Personnel Assessment, Report prepared for the Commonwealth of Pennsylvania, December 1978.
31. Skinner, H.A., Jackson, D.N., and Rampton, G.M., "The Personality Research in a Canadian Context: Does Language Make a Difference?" Canadian Journal of Behavioral Sciences, Vol. 8, pp. 156-168, 1976.
32. Joaquin, J.B., The Personality Research Form and its Utility in Predicting Undergraduate Pilot Training Performance in the Canadian Forces, Working Paper 80-12, Canadian Forces Personnel Applied Research Unit, Willowdale, Ontario, 1980.
33. Bucky, S.F. and Ridley, S.L., "California Personality Inventory as a Predictor of Success in the Naval Flight Program." Aerospace Medicine, Vol. 43, pp. 971-973, 1972.
34. Weider, A., Wolff, H.G., Brodman, K., Mittelmann, B., and Wechsler, D., Cornell Index, Psychological Corporation, New York, NY, 1945.
35. Buros, O.K., The Sixth Mental Measurements Yearbook, Gryphon Press, Highland Park, NJ, 1965.
36. Barry, J.R., Sells, S.B., and Trites, D.K., Psychiatric Screening of Flying Personnel: Research on the Cornell Word Form, TR-7, School of Aviation Medicine, Randolph AFB, TX, 1954.
37. Trites, D.K. and Kubala, A.L., "Characteristics of Successful Pilots." Journal of Aviation Medicine, Vol. 28, pp. 34-40, 1957.
38. Fleischman, H.L., Ambler, R.K., Peterson, F.E., and Lane, N.E., The Relationship of Five Personality Scales to Success in Naval Aviation Training, TR-968, Naval Aerospace Medical Institute, Naval Air Station, Pensacola, FL, 1966.
39. Bucky, S.F. and Spielberger, C.D., "State and Trait Anxiety in Voluntary Withdrawal of Student Naval Aviators from Flight Training." Psychological Reports, Vol. 33, pp. 351-354, 1973.
40. Krahenbuhl, G.S., Marett, J.R., and Kinz, N.W., Stress and Performance in T-37 Pilot Training, Air Force Human Resources Laboratory, TR-77-3, Williams Air Force Base, AZ, May 1977.
41. Cattell, R.B., Eber, H.W., and Tatsuoka, M.M., Handbook for the Sixteen Personality Factor Questionnaire, Institute for Personality and Ability Testing, Champaign, IL, 1970.
42. Bartram, D., Development and Evaluation of Micropat Version 4, TR-184, Ministry of Defence, London, and Ergonomics Research Group, University of Hull, U.K., 1986.

43. Luk'yanova, N.F., Personality Characteristics of Pilot-cadets with Different Marks in Flight Disciplines, U.S. Army Foreign Science and Technology Center, Charlottesville, VA, 1977.
44. Cattell, R.B. and Stice, G.F., Handbook for the Sixteen Personality Factor Questionnaire, Institute for Personality and Ability Testing, Champaign, IL, 1962.
45. Bartram, D. and Dale, H.C.A., Micropat Version 3: A Description of the Fully Automated Personnel Selection Testing System being Developed for the Army Air Corps, ERG/Y6536/83/7, Ergonomics Research Group Report University of Hull, U.K., 1983.
46. Heist, P. and Yonge, G., Omnibus Personality Inventory Manual Form F, Psychological Corporation, New York, NY, 1969.
47. Rossman, J.E. and Kirk, B.A., "Factors Related to Persistence and Withdrawal Among University Students." Journal of Counseling Psychology, Vol. 17, pp. 56-62, 1970.
48. Griffin, R.G. and Hopson, J.A., An Evaluation of the Omnibus Personality Inventory in the Prediction of Attrition in Naval Aviation Training, TR-1253, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1978.
49. Kennedy, R.S. The Relationship between Habituation to Vestibular Stimulation and Vigilance: Individual Differences and Subsidiary Problems, Monograph 20, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1972.
50. Kennedy, R.S., "Motion Sickness Questionnaire and Field Independence Scores as Predictors of Success in Naval Aviation Training." Aviation, Space, and Environmental Medicine, Vol. 46, pp. 1349-1352, 1975.
51. Cullen, J.F., Harper, C.R., and Kidera, G.J., "Perceptual Style Differences between Airline Pilots and Engineers." Aerospace Medicine, Vol. 40, pp. 407-408, 1969.
52. Witkin, H.A., "Individual Differences in the Ease of Perception of Embedded Figures." Journal of Personality, Vol. 19, pp. 1-16, 1950.
53. Witkin, H.A., Lewis, R.B., Hertzman, M., Machover, K., Meissner, P.B., and Wapner, S., Personality through Perception, Harper's Publishing, New York, NY, 1954.
54. Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., and Karp, S.A., Psychological Differentiation: Studies of Development, Wiley Publishing, New York, NY, 1962.
55. Witkin, H.A., Oltman, P.K., Raskin, E., and Karp, S.A., A Manual for the Embedded Figures Test, Consulting Psychologists Press, Palo Alto, CA, 1971.

56. Barrett, G.V. and Thornton, C.L., "Cognitive Style Differences between Engineers and College Students." Perceptual and Motor Skills, Vol. 25, p. 789, 1967.
57. Gardner, R.W., Jackson, D.N., and Messick, S.J., "Personality Organization in Cognitive Controls and Intellectual Abilities." Psychological Issues, Vol. 2, Whole No. 8, 1960.
58. Goldstein, A.G. and Chance, J.E., "Effects of Practice on Sex-related Differences in Performance on Embedded Figures." Psychonomic Science, Vol. 3, pp. 361-362, 1965.
59. Hyde, J.S., Geiringer, E.R., and Yen, W.M., "On the Empirical Relation between Spatial Ability and Sex Differences in other Aspects of Cognitive Performance." Multivariate Behavioral Research, Vol. 10, pp. 289-309, 1975.
60. Johnson, S., Flinn, J.M., and Tyler, Z.E., "Effect of Practice and Training in Spatial Skills on Embedded Figures Scores of Males and Females." Perceptual and Motor Skills, Vol. 48, pp. 975-984, 1979.
61. Long, G.M., Field Dependency-Independancy: A Review of the Literature, Monograph 19, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1972.
62. McGee, M.G., "Human Spatial Abilities: Psychometric Studies and Environmental, Genetic, Hormonal, and Neurological Influences." Psychological Bulletin, Vol. 86, pp. 889-918, 1979.
63. Sherman, J.A., "Field Articulation, Sex, Spatial Visualization, Dependency, Practice, Laterality of the Brain and Birth Order." Perceptual and Motor Skills, Vol. 38, pp. 1223-1235, 1974.
64. Windsor, A., "Sex and Profession as Determinants of Field Dependence." Perceptual and Motor Skills, Vol. 57, pp. 617-618, 1983.
65. Jackson, D.N., Messick, S., and Myers, C.T., The Role of Memory and Color in Group and Individual Embedded Figures Measures of Field Independence, Research Bulletin of Educational Testing Service, Princeton, NJ, pp. 34-62, 1962.
66. Edwards, A.L., Manual for the Edwards Personal Preference Schedule, Psychological Corporation, New York, NY, 1959.
67. Fry, G.E. and Reinhardt, R.F., "Personality Characteristics of Jet Pilots as Measured by the Edwards Personal Preference Schedule." Aerospace Medicine, Vol. 40, pp. 484-486, 1969.
68. Novello, J.R. and Youssef, I.Z., "Psycho-social Studies in General Aviation: II: Personality Profile of Female Pilots." Aerospace Medicine, Vol. 45, No. 6, pp. 630-633, 1974.
69. Peterson, F.E., Lane, N.E., and Kennedy, R.S., The Relationship of the Edwards Personal Schedule to Success in Naval Flight Training, TR-46, Naval Aerospace Medical Institute, Pensacola, FL, 1965.

70. Fine, P.M. and Hartman, B.O., Psychiatric Strengths and Weaknesses of Typical Air Force Pilots, USAF School of Aerospace Medicine, TR-68121, Brooks Air Force Base, TX, November 1968.
71. Jennings, C.L., "The Use of Normative Data in the Psychological Evaluation of Flying Personnel." International Psychiatry Clinics, Vol. 4, Whole No. 1, 1967.
72. Novello, J.R and Yousset, I.Z., "Psycho-social Studies in General Aviation: I: Personality Profile of Male Pilots." Aerospace Medicine, Vol. 45, No. 2, pp. 185-188, 1974.
73. Reinhardt, R.F., "The Compulsive Flyer." Aerospace Medicine, Vol. 37, pp. 411-413, 1966.
74. Ashman, A., and Tefler, R., "Personality Profiles of Pilots." Aviation, Space, and Environmental Medicine, Vol. 10, pp. 940-943, 1983.
75. Robertson, D.W., "Prediction of Naval Aviator Carrier Motivation and Job Satisfaction from the Strong Vocational Interest Blank." Dissertation Abstracts International, Vol. 35, p. 4244, 1975.
76. Guinn, N , Bart, M.V., and Leisey, S.A., Background and Interest Measures as Predictors of Success in Undergraduate Pilot Training, AFHRL-TR-76-9, Air Force Human Resources Laboratory, Lackland Air Force Base, TX, May 1976.
77. Doll, R.E., Ambler, R.K., Lane, N.E., and Bale, R.M., "Vocational Interest Differences between Students Completing the Naval Aviation Training Program and Students Voluntarily Withdrawing." Proceedings of the 80th Annual Meeting, American Psychological Association, 1972.
78. Jenkins, C.D., Zyzanski, S.J., and Rosenman, R.H., Jenkins Activity Survey, Psychological Corp., New York, NY, 1979.
79. Spence, J.T., Helmreich, R.L., and Pred, R.S., "Impatience Versus Achievement Strivings in the Type A Pattern: Differential Effects on Students' Health and Academic Achievement." Journal of Applied Psychology, Vol. 72, pp. 522-528, 1987.
80. Helmreich, R.L. and Spence, J.T., "The Work and Family Orientation Questionnaire: An Objective Instrument to Assess Components of Achievement Motivation and Attitudes toward Family and Career." SAS: Catalog of Selected Documents in Psychology, Vol. 8, MS 1677, 1978.
81. Rotter, J.B., "Generalized Expectancies for Internal Versus External Control of Reinforcement." Psychological Monographs: General and Applied, Vol. 80. pp. 1-28, 1966.
82. Wichman, H. and Ball, J., "Locus of Control, Self-serving Biases, and Attitudes towards Safety in General Aviation Pilots." Aviation, Space, and Environmental Medicine, Vol. 54, pp. 507-510, 1983.

83. Helmreich, R.L., Sawin, L.L., and Carsrud, A.L., "The Honeymoon Effect in Job Performance: Temporal Increases in the Predictive Power of Achievement Motivation." Journal of Applied Psychology, Vol. 71, pp. 185-188, 1986.
84. Dolgin, D.L., Gibb, G.D., Nontasak, T., and Helm, W.R., Instructor Pilot Evaluations of Key Naval Primary Flight Training Criteria, NAMRL-1331, Naval Aerospace Medical Research Laboratory, Pensacola, FL, June 1987.
85. Shannon, R.H. and Waag, W.L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, TR-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1972.
86. Melton, A.W., Apparatus Tests, Army Air Forces Aviation Psychology Program Research Report No. 4. U.S. Government Printing Office, Washington, DC, 1947.
87. Helmreich, R.L., Pilot Selection and Training, Paper presented at the American Psychological Association Annual Meeting, Washington, DC, 1982.
88. Spence, J.T., Helmreich, R.L., and Holahan, C.K., "Negative and Positive Components of Psychological Masculinity and Femininity and their Relationships to Self-reports of Neurotic and Acting Out Behaviors." Journal of Personality and Social Psychology, Vol. 37, pp. 1673-1682, 1979.
89. Green, R.F., "Does a Selection Situation Induce Testees to Bias their Answers on Interest and Temperment Tests." Educational and Psychological Measurement, Vol. 1951, pp. 503-515, 1951.
90. Stricker, L.J., "Test-wiseness on Personality Scales." Journal of Applied Psychology Monograph, Vol. 53, pp. 1-18, 1969.
91. Wiener, S., Officer Candidate Tests, Arco Publishing Company, New York, NY, 1985.
92. Bartram, D. and Bayliss, R., "Automated Testing: Past, Present and Future." Journal of Occupational Psychology, Vol. 57, pp. 221-237, 1984.
93. Kantor, J. and Bordelon, V., "The USAF Pilot Selection and Classification Research Program." Aviation, Space, and Environmental Medicine, Vol. 56, pp. 254-257, 1985.
94. Mullins, C., Two Measures of Decisiveness, TM-62-5, Personnel Research Laboratory, Lackland Air Force Base, San Antonio, TX, 1962.
95. Siem, F.M., Carretta, T.R., and Mercatante, T.A., Personality, Attitudes and Pilot Training Performance, AFHRL TP-87-36, Air Force Human Resources Laboratory, Brooks Air Force Base, TX, 1987.

96. Carretta, T.R., "The Basic Attributes Tests (BAT) System: A Preliminary Evaluation of Three Cognitive Subtasks." Proceedings of the Human Factors Society 30th Annual Meeting, pp. 1321-1325, 1986.
97. Gibb, G.D. and Dolgin, D.L., "Validation of a Computer-Based Aviation Secondary Selection System For Student Naval Aviators." Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 807-811, Anaheim, CA, 1988.
98. Lambirth, T.T., Gibb, G.D., and Alcorn, J.D., "Use of a Behavior-based Personality Instrument in Aviation Selection." Educational and Psychological Measurement, Vol. 46, pp. 973-978, 1986.
99. Long, G.M. and Shelnutt, J.B., Decisional Differences among Individuals: A Signal Detection Theory Approach, NAMRL-1185, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1973.
100. Slovic, P., "Manipulating the Attractiveness of a Gamble without Changing its Expected Value." Journal of Experimental Psychology, Vol. 79, pp. 139-145, 1969.
101. Atkinson, J.W. and Litwin, G.H., "Achievement Motive and Test Anxiety Conceived as Motive to Approach Success and Motive to Avoid Failure." Journal of Abnormal and Social Psychology, Vol. 60, pp. 52-63, 1960.
102. Pruitt, D.G. and Teger, A.I., "The Risky Shift in Group Betting." Journal of Experimental Social Psychology, Vol. 5, pp. 115-126, 1969.
103. Atkinson, J.W., Bastian, J.R., Earl, R.W., and Litwin, G.H., "The Achievement Motive, Goal Setting, and Probability Preferences." Journal of Abnormal and Social Psychology, Vol. 60, pp. 27-36, 1960.
104. Gilson, C.R., Individual Differences in Risk-taking, Technical Report, No. 13, Office of Naval Research, June 1968.
105. Kogan, N. and Wallach, M.A. (Eds.), Risk Taking: A Study in Cognition and Personality, Holt, Rinehart, and Winston, New York, NY, 1964.
106. Imhoff, D.L. and Levine, J.M., Perceptual-motor and Cognitive Performance Task Battery for Pilot Selection, AFHRL-TR-80-27, Air Force Human Resource Laboratory, Brooks AFB, TX, 1981.
107. O'Conner, M.F., Peterson, C.R., and Palmer, T.J., "Stakes and Probabilities in Information Purchase." Organizational Behavior and Human Performance, Vol. 7, pp. 43-52, 1972.
108. Stoner, J.A., "Risky and Cautious Shifts in Group Decisions: The Influence of Widely Held Values." Journal of Experimental Social Psychology, Vol. 4, pp. 442-459, 1968.

109. Shull, R.N., Dolgin, D.L., and Gibb, G.D., The Relationship between Flight Training Performance, a Risk Assessment Test and the Jenkins Activity Survey, NAMRL-1339, Naval Aerospace Medical Research Laboratory, Pensacola, FL, June 1988.
110. Kragh, V., "The Defense Mechanism Test: A New Method for Diagnosis and Personnel Selection." Journal of Applied Psychology, Vol. 44, pp. 303-309, 1960.
111. Neuman, T., Cross-validation of the Defense Mechanism Test by the Criterion of Passing or Failure in Basic Military Flight Training, FTD-HC-23-1150-72, Foreign Technology Division, Wright-Patterson AFB, Ohio, August 1972, (AD-904-138L).
112. Neuman, T., Influence of DMT on Economy of Training and Flight Safety in the Royal Swedish Airforce 1970/71-1985/86, The Medical College of Saint Bartholomew's Hospital, London, December 1986.
113. Sjöberg, L., "The Value of DMT in the Selection of Pilots." Nordisk Psykologi, Vol. 33, pp. 241-248, 1981.
114. Vaernes, R.J., "The Defense Mechanism Test Predicts Inadequate Performance Under Stress." Scandinavian Journal of Psychology, Vol. 23, pp. 37-43, 1982.
115. Kolb, L.C., Modern Clinical Psychiatry (8th Ed.), W.B. Saunders Company, Philadelphia, PA, 1973.
116. Stoker, P., "An Empirical Investigation of the Predictive Validity of the Defence Mechanism Test in the Screening of Fast-jet Pilots for the Royal Air Force." Projective Psychology, Vol. 27, pp. 7-12, 1982.
117. Alkov, R.A., Gaynor, J.A., and Borowsky, M.S., "Pilot Error as a Symptom of Inadequate Stress Coping." Aviation, Space, and Environmental Medicine, Vol. 56, pp. 244-247, 1985.
118. McGuire, F.L., "Personality Factors in Highway Accidents." Human Factors, Vol. 18, pp. 433-436, 1976.
119. Alkov, R.A., Borowsky, M.S., and Gaynor, J.A., "Stress Coping and the U.S. Navy Aircrew Factor Mishap." Aviation, Space, and Environmental Medicine, Vol. 53, pp. 1112-1115, 1982.
120. Jenson, R.S. and Benel, R.A., Judgment Evaluation and Instruction in Civil Pilot Training, TR-RD-78-24, Federal Aviation Administration Springfield, VA, 1977.
121. Lester, L.F. and Bombaci, D.H., "The Relationship between Personality and Irrational Judgment in Civil Pilots." Human Factors, Vol. 26, pp. 565-572, 1984.

Other Related NAMRL Publications

Other related NAMRL publications are cited in the text and listed in the references.